Many queries are submitted to search engines by **right-clicking** some selected words in Web browser.

- Right-click queries are currently processed in the same way by a search engine as queries submitted through other means.
- The source document from which the query is marked for search provides sufficient contextual information to determine the right semantic of the query.

### Context-Aware Search Framework

- Contextual information can be extracted to enable **Context-Aware Search** for better user search experiences.
- **Two main research questions:**
  - Given the source document of a right-click query, which component of the document (e.g., title, full text, paragraph containing the right-click query) is best in providing contextual information for the query?
  - What contextual information (e.g., words, nouns, or noun phrases) shall be extracted to augment the query?

### Weighting Scheme

- **Frequency-based weighting:** TF · IDF weighting scheme
- **Proximity-based weighting:** A term is more important if (i) its TF · IDF score is large, (ii) it occurs for multiple times in the selected text component, and (iii) the occurrences are close to the query in terms of proximity distance.

\[
p_w(t_i) = \frac{\sum_{j=1}^{k} f_w(t_i)}{\text{dist}(V, q)}
\]

- **Phrase weighting:** (i) the phrase’s TF · IDF score \( f_w(s) \) by treating each phrase as a token, and (ii) the average frequency of all terms contained in the phrase \( \sum_{i \in s} f_i / |s| \), where \( |s| \) is the number of terms in phrase \( s \).

\[
s_w(s) = f_w(s) \frac{\sum_{i \in s} f_i}{|s|}
\]

### Evaluation

- **Data collection:** A user study using 20 news articles from Yahoo! News selected mainly based on two criteria: (i) article contains an ambiguous query term; (ii) two or more articles contain the same query term but with different semantics.
- **Baseline method:** Top-8 keywords recommended by Google search engine for the right-click query.
- **Our methods:** Top-8 keywords (or at most 8 words from top-\( N \) phrases) ranked by the text component \( T_s \) and feature extraction scheme \( F_s \).
- **Evaluation metric:** Each of the top-8 ranked words is manually judged to be relevant or irrelevant based on the context of the news article and the right-click query. Precision is used to evaluate the methods.

### Context Extraction

- **Seven text components** \( T_1 \) – \( T_7 \):
  - \( T_1 \) Full text of the page
  - \( T_2 \) Paragraph of the selected query word(s)
  - \( T_3 \) Title of the page
  - \( T_4 \) Title, the first and last paragraph
  - \( T_5 \) Paragraphs containing the query word(s)
  - \( T_6 \) Meta description and keyword of the page
  - \( T_7 \) Full text of the current and referenced articles

- **Five feature extraction schemes** \( F_1 \)–\( F_5 \):
  - \( F_1 \) Words
  - \( F_2 \) Words
  - \( F_3 \) Nouns
  - \( F_4 \) Nouns
  - \( F_5 \) Noun Phrases

### Results

**Observations:**

- Noun phrases with phrase weighting (\( F_5 \)) is the best context feature extraction scheme
- Paragraphs containing the query words (\( T_5 \)) are the best text components for query context extraction
- Proximity-based weighting scheme adversely affects the precision compared with frequency-based weighting scheme.

Between nouns and any words, using the same weighting schemes, nouns define better contextual information than any words.